A simple effect size estimator for single case designs using WinBUGS

David Rindskopf
William Shadish
Larry V. Hedges

City University of New York Graduate Center
University of California, Merced
Northwestern University

March 13, 2012

Thanks to
Institute of Education Sciences
for support through grants:
R305D100046
R305D100033

Usual Effect Size for Two Group Study

- Most common effect size (ES) measures compare means
- Because different studies use different scales, we must adjust for variance
- Thus we have the basic form ES $= \frac{\overline{X}_T - \overline{X}_C}{s}$
- Notation: T = Treatment group, C = control group, s is standard deviation (of control or pooled)
- Several adjustments have been proposed to reduce bias

Extension to Single Case Design (SCD)

- Why should we? After all, the dependent variable is often already on a scale that is directly interpretable.
- Because: Sometimes you want to compare SCDs with between-group studies that use different outcomes
- One approach: divide (treatment - baseline) by respondent's baseline standard deviation
- Why not? This is a within-person standard deviation
- Between-person variation (from two-group studies) is usually much larger, thus not comparable

An Effect Size (ES) Comparable to Usual d

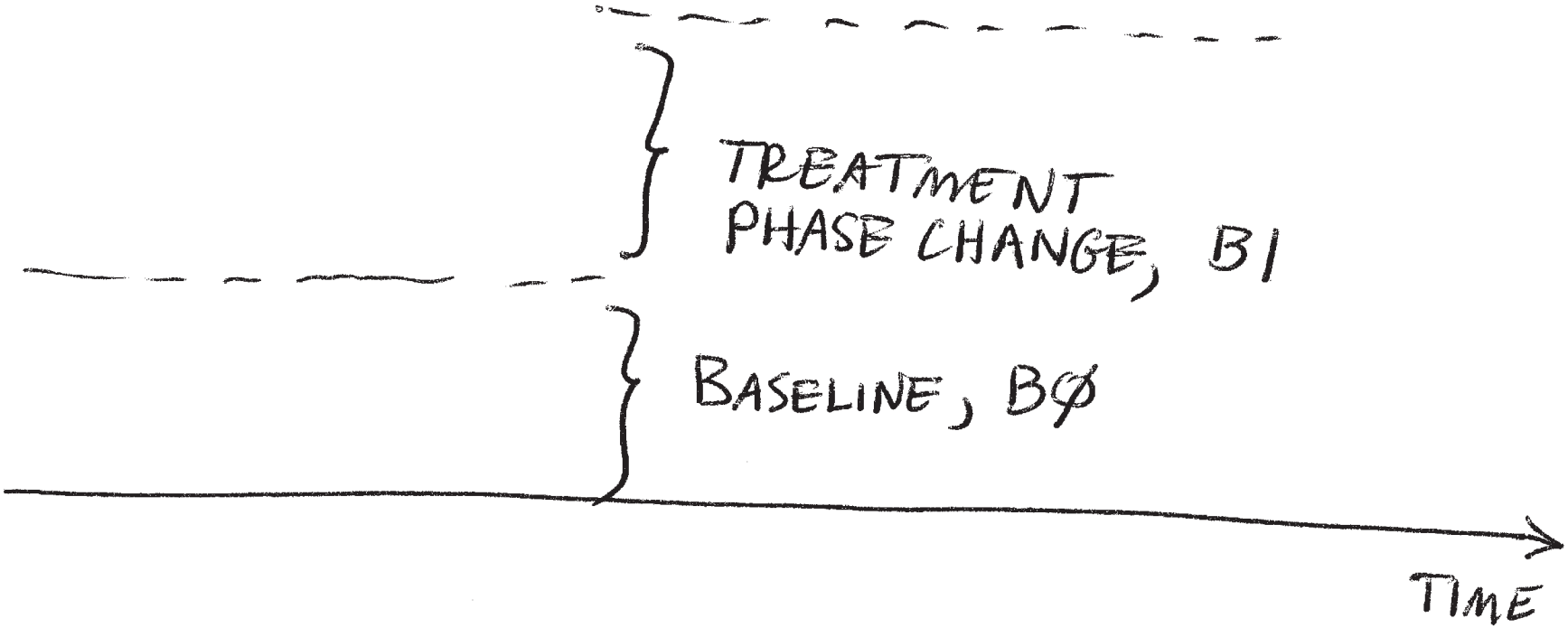
- Suppose an SCD has several participants
- Use variation among them as denominator
- Several options are possible, each with advantages and disadvantages
- Method 1: Divide by estimate of between-person variance (flawed, though it seems natural)
- Method 2: Divide by between + within variance (right, though not intuitive)

Complication: Dependent variables are often counts, not continuous

- Ex 1: How many times does S hit another child?
  - No (theoretical) max
  - Poisson (or more complicated) distribution
  - Stat notation: $Y \sim Poisson(\lambda)$
  - BUGS (Bayesian software) notation:
    `y[i] ~ dpois(lambda[i])`
- Ex 2: How many HW probs (out of 10) does S attempt?
  - attempt/10 = proportion
  - binomial (or more complicated) distribution
  - Stat notation: $Y \sim Binomial(p, n)$
  - BUGS notation: `y[i] ~ dbin(p,n)`

Graphical representation of important concepts

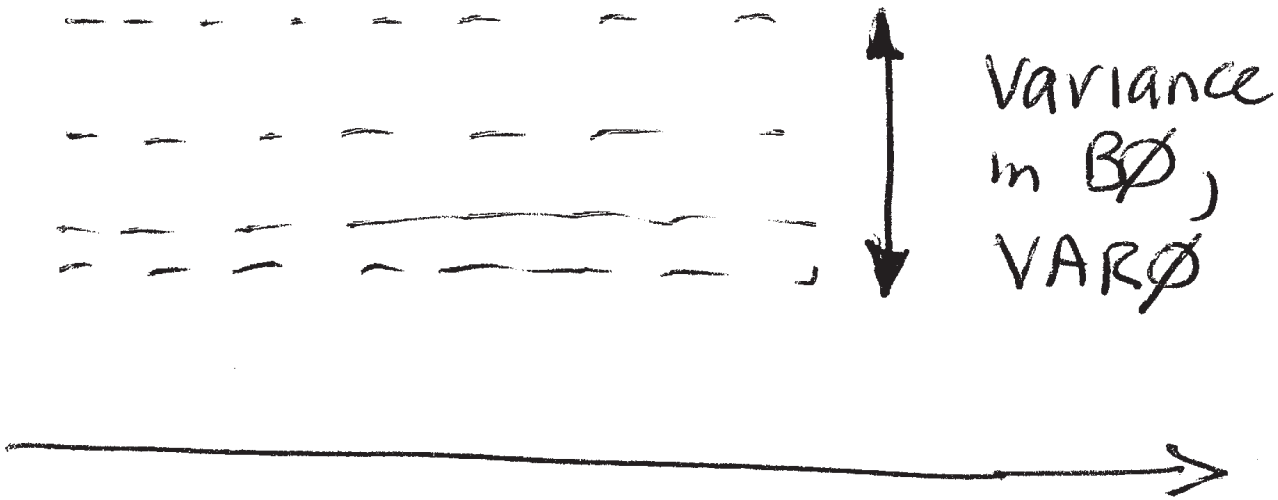
(1) Sketch of simple phase-change model for a single subject



Each subject has an average baseline value, and an average change during the treatment phase.

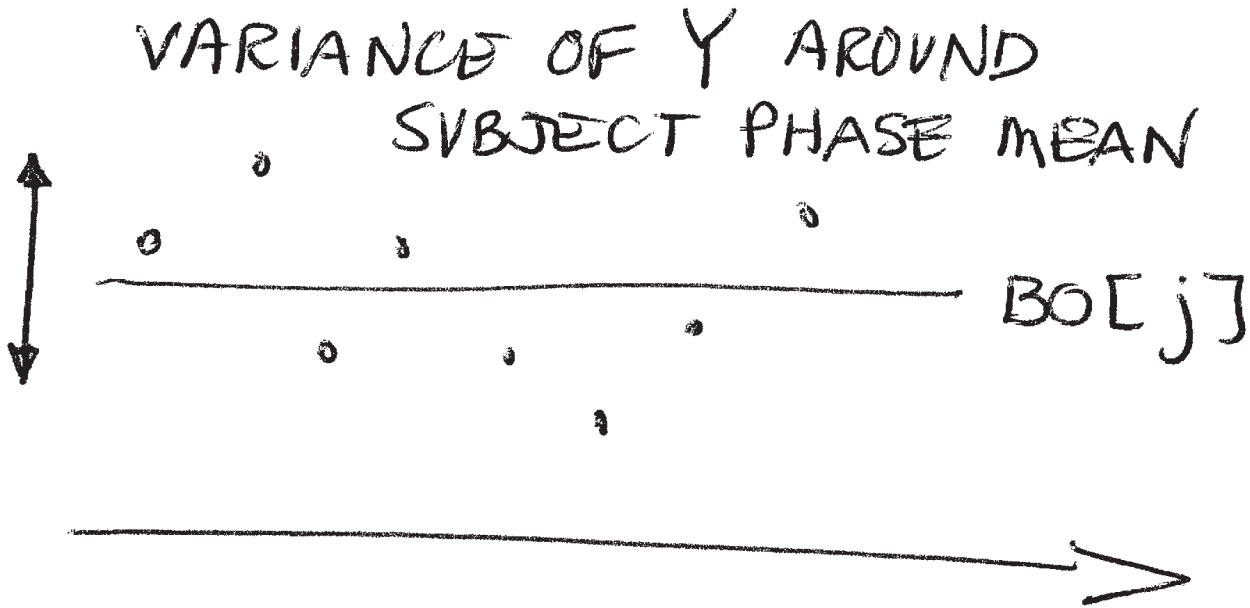
Grahical representation

(2)Subjects vary in their baseline levels



Graphical representation

(3) Within each subject, there is variation around the phase mean



Statistical model for a continuous outcome and one person-level predictor

Model for individual observations for subject j at time i:

$$Y_{ij} = \beta_{0j} + \beta_{1j} Phase_{ij} + r_{ij}$$

$$Phase_{ij} = \begin{cases} 0 & \texttt{for baseline} \\ 1 & \texttt{for treatment} \end{cases}$$

We can model the baseline value of subject j as a function of Sex:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} Female_j + u_{0j}$$

(where $Female_j$ is 0 for males, 1 for females)

And also the treatment effect of subject j as a function of Sex:

$$\beta_{1j} = \gamma_{10} + \gamma_{11} Female_j + u_{1j}$$

Simplified and combined statistical model for a continuous outcome (no person-level predictor)

Model for individual observations for subject j at time i:

$$Y_{ij} = \beta_{0j} + \beta_{1j} Phase_{ij} + r_{ij}$$

Model for baseline value of subject j:

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

Model for treatment effect of subject j:

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

Combined model, substituting into first equation:

$$Y_{ij} = (\gamma_{00} + u_{0j}) + (\gamma_{10} + u_{1j}) Phase_{ij} + r_{ij}$$

We often use the combined model for certain software packages, including the BUGS software I will be using.

More details on model

- $Var(u_{0j}) = \tau_{00}$ is part of the between-person variation
- $Var(r_{ij}) = \sigma^2$ is within person variation
  - contributes to usual denominator in ES
  - can't be separated from between-person variation without repeated measurements of the same person
- Thus, we might naively use $ES_1 = \gamma_{10}/\sqrt{\tau_{00}}$
- But more properly we would define $ES_2 = \gamma_{10}/\sqrt{\tau_{00} + \sigma^2}$
- On logit scale, $\sigma^2 \approx 1/(n\pi(1-\pi))$

Logistic version of model

- Our data has number of successes in 10 trials, measured each day (or session) for the period of the study
- We need the equivalent of a logistic regression
- Outcome is the logarithm of the odds, called the logit
- In GLM (generalized linear models) this is expressed by separating the linear part of the model from the (logit) transform of the dependent variable

Logistic model, cont.

We represent the combined model as follows:

$\eta_{ij} = (\gamma_{00} + u_{0j}) + (\gamma_{10} + u_{1j})Phase_{ij}$

$\eta_{ij} = \ln(\frac{\pi_{ij}}{1-\pi_{ij}}) = logit(\pi)$

$Y_{ij} \sim binomial(\pi_{ij}, n_{ij})$

Where Y_{ij} is the count of events out of n_{ij} trials,
each with probability π_{ij} of the event occuring
(Often $n_{ij} = n$; that is, it is constant across time and subjects)

Why Bayesian?

- Bayesian philosophy: Statistics is about using data to revise beliefs about unknown values (parameters)
- Initial (prior) beliefs could be vague (noninformative) or based on evidence available before collecting current data
- Information in the data is combined with any prior information to produce a Posterior Distribution, which summarizes our beliefs after seeing the data
- Bayesian models resemble usual models, except for specification of priors and interpretation of outcome
- Bayesian interpretation of interval: Probability is .95 that the parameter is in the interval (natural, but wrong in classical stat)

Why (Win)BUGS?

- Some Bayesian computational methods, including those used in BUGS, allow simple ways to make inferences about derived quantities
- In particular, we want to make inferences about effect sizes, which are complicated quantities
- BUGS will make it (relatively) simple for us to make inferences about effect sizes
- The following slides will show how to write a simple logistic regression model, compute ES, and interpret results

Data Structure

```
subj[ ]  r[ ]  phase[ ]
1   4    0
1   5    0
1   5    0
1   4    0
1   5    0
1  10    1
1   8    1
....
1   9    1
1   9    1
2   2    0
2   3    0
...
```

```
2  10    1
2   9    1
3   6    0
3   6    0
...
3   9    1
3   9    1
4   4    0
4   3    0
...
4   9    1
4   9    1
END DATA
```

WinBUGS Code 1: Basic Model

First some comments so I remember what I'm doing:

```
#  binomial, 10 trials per session
#  4 respondents, 2 phases (AB), multiple baseline
#  p(yes) goes from about .5 in baseline to .8 or .9
```

Next the model, first expressing the logit (log-odds) as a function of phase, then the distribution as a binomial:

```
model
{ for (i in 1:103)
    logit(p[i]) <- base[subj[i]] +
        trt[subj[i]] * phase[i]
    r[i] ~ dbin(p[i],10) }
```

Each subject has his/her own baseline and treatment effect, with mean mu and precision (1/variance) prec:

```
  for (j in 1:4)
   { base[j] ~ dnorm(mu0, prec0)
     trt[j]  ~ dnorm(mu1, prec1) }
```

WinBUGS Code 2: Priors and Create Variances, SDs

First, relatively uninformative priors for means and standard deviations of baseline and treatment effect:

```
mu0 ~ dnorm(0,.001)
mu1 ~ dnorm(0,.001)
prec0 ~ dgamma(.01,.01)
prec1 ~ dgamma(.01,.01)
```

Next, define variances and standard deviations, because it's hard for most of us to think in terms of precisions:

```
var0 <- 1/prec0
var1 <- 1/prec1
sd0 <- sqrt(var0)
sd1 <- sqrt(var1)
```

WinBUGS Code 3: Create New Variables

Find estimate of within-person variation at baseline;
must first transform from logit scale to find mean probability:

```
odds0 <- exp(mu0)
prob0 <- odds0/(1+odds0)
sigma.2 <- 1/(10 * prob0 * (1-prob0))
```

Next create total variance (between + within), and denominator for effect size estimate:

```
var.tot <- sigma.2 + var0
sd.tot  <- sqrt(var.tot)
```

Compute ES estimates, first wrong and then correctly:

```
# Uses only variation in average baselines:
  es.bet.1  <- mu1/sd0


# (Properly) uses total variation:
  es.bet.2  <- mu1/sd.tot
}
```

Output: Effect Sizes 1

node	mean	sd	2.5%	median	97.5%
es.bet.1	5.5219	3.52	1.3147	4.8045	14.353
es.bet.2	2.7805	0.74798	1.195	2.8365	4.119

- First has a denominator that is too small, and therefore the ES estimate is too large
- Second should be (at least approximately) right

Advantages of MCMC(BUGS): Not just estimate, but also

- Standard error (called sd in output)
- CI (2.5% – 97.5%)
- Info on skewness of distribution

Output: Effect Sizes 2

Rounded estimates for ES2 (correct estimate):

node	mean	sd	2.5%	median	97.5%
es.bet.2	2.78	0.75	1.20	2.84	4.12

For ES2, we have

- Evidence of minor skewness:
  Compare $2.84 - 1.20 = 1.64$ to $4.12 - 2.84 = 1.28$
- More evidence skew is minor:
  Mean (2.78) is close to median (2.84)
- Empirical 95 percent credible (confidence) interval:
  (1.20, 4.12)
- Thus ES could be as small as about 1, or as large as about 4
- Quite wide interval due to small number of respondents

Output : Basic Parameters

node	mean	sd	2.5%	median	97.5%
mu0	-0.099319	0.38842	-0.84193	-0.10322	0.66174
mu1	2.3425	0.36558	1.6788	2.3369	3.0548
sd0	0.59467	0.44389	0.17013	0.48249	1.7053
sd1	0.47212	0.43196	0.089285	0.36225	1.5825

- Baseline average odds : $exp(-.099) = .905$
- Baseline average proportion: $.905/1.905 = .475$
- Treatment phase average odds : $exp(-.909 + 2.343) = 9.42$
- Treatment phase average proportion : $9.42/10.42 = .904$

Output : Variances

Variances among subjects in baseline log-odds
and treatment effects:

node	mean	sd	2.5%	median	97.5%
var0	0.55067	1.7968	0.028945	0.2328	2.9079
var1	0.40949	1.4639	0.007972	0.13123	2.5044

Variance within phases for subjects (sigma.2) and total (var.tot):

sigma.2	0.42011	0.15086	0.40001	0.40405	0.50187
var.tot	0.97079	1.9037	0.43176	0.64223	3.3746
sd.tot	0.91024	0.37715	0.65708	0.80139	1.837

Estimate for Individuals

Baseline log-odds for each person:

node	mean	sd	2.5%	median	97.5%
base[1]	0.032035	0.18773	-0.33265	0.032001	0.40112
base[2]	-0.39729	0.20606	-0.81881	-0.39097	-0.01188
base[3]	0.40382	0.21969	-0.02369	0.40348	0.83808
base[4]	-0.44905	0.20319	-0.84266	-0.44938	-0.04398

Treatment effect (on log-odds scale) for each person:

trt[1]	2.5287	0.30277	1.9984	2.5041	3.1901
trt[2]	2.5158	0.2903	1.9742	2.5042	3.1398
trt[3]	2.3207	0.30109	1.7497	2.3133	2.9469
trt[4]	1.9931	0.27734	1.4186	2.006	2.4951

Problems with this approach

- Standardized measures aren't in original scale
- With small number of observations in any phase, the difference between phases is not well-estimated
- With small number of respondents, standard deviation among respondents is not well-estimated
- If respondents are selected for low (or high) initial status, between-respondent variation may be artificially low compared to between-group studies (needs checking, at least)

Conclusions

- Bayesian computation allows relatively simple
  - Estimation of effect size
  - Production of confidence interval (CI)
  - Estimates for each individual, including CI
- Some additional training (beyond HLM) is needed to set up the model, do computations, and interpret results
- This procedure should be useful for SCD researchers who want to produce results comparable to between-group studies

References

N Best, D Spiegelhalter, A Thomas, and C Brayne. (1996) Bayesian analysis of realistically complex models, *Journal of the Royal Statistical Society: Series A*, *159* , 323-342.

D Lunn, A Thomas, N Best, and D Spiegelhalter (2000) WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility.*Statistics and Computing*, *10* , 325-337.

D Spiegelhalter, A Thomas, N Best, and W Gilks, *BUGS: Bayesian inference using Gibbs sampling, Version 0.50*, MRC Biostatistics Unit, Cambridge, UK, 1995.

Spiegelhalter, D., Thomas, A. and Best, N., (2000). *WinBugs Version 1.4 User Manual*.